知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum
E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

国内外关联数据研究热点对比分析

◎ 王振蒙 [1,2]

[1] 中国科学院大学 北京 100049
[2] 中国科学院成都文献情报中心 成都 610041

摘要：[**目的 / 意义**] 国内关联数据研究尚出于研究发展的初期，理论研究和应用研究也都尚未成熟。通过对国内外研究热点主题对比分析，为国内关联数据研究起提供借鉴。[**方法 / 过程**] 通过数据分析和文献调研，分别从发文量、研究群体、词频以及主题识别等方面进行中外研究的对比分析，从中发现国内外关联数据研究存在的差距。[**结果 / 结论**] 在对比分析的基础上，针对国内关联数据研究提出研究本土化、注重应用型研究、优化研究群体结构以及关注关联数据学科化应用 4 点发展建议。

关键词：关联数据 研究热点 对比分析 数据分析
分类号：G250.7

引用格式：王振蒙. 国内外关联数据研究热点对比分析 [J/OL]. 知识管理论坛 , 2016, 1(3): 163-173[引用日期]. http://www.kmf.ac.cn/paperView?id=33.

1 引言

自 2006 年 T. Berners-Lee 提出关联数据的概念以来，研究人员便开始对关联数据进行深入的研究。关联数据是 W3C 推荐的一种用于发布和联接各类数据的规范，旨在建立一个映射所有自然、社会和精神世界的数据网络，通过对大千世界万事万物及其相互之间关系进行机器可读的描述，来构建一个一个富含语义的、互联互通的知识海洋，使人们可以在更大范围内准确、高效、可靠地查找、分享、利用这些相互关联的信息和知识 [1]。国内对关联数据的关注最早始于 2008 年，由曾蕾、刘炜等在于上海举办的 " 数字环境下图书馆前沿问题 " 研讨班上将关联数据这一概念正式引入国内。自此，国内关于关联数据的研究不仅在数量上日渐增多，研究的主题也日趋广泛。本文通过对国内外相关研究文献进行共词分析、主题挖掘，并借助 SPSS、Sati 等工具进行统计、主题聚类分析等，试图通过对比找出国内关联数据研究的特点与不足，为国内关联数据研究及发展提供指导。

2 数据来源与研究方法

本文研究数据力求精准、全面，在中文数据的收集过程中，笔者发现除核心词汇 " 关联数据 " 以外还有一些其他关键词，笔者也将它们列入检索式；在保证数据全面性的同时，考虑到中文词汇的多义性，还对初步的检索结果做了进一步的人工筛选，最终得到 381 条文献数据。在中文数据的筛选过程中，笔者发现 381

作者简介：王振蒙 (ORCID:0000-0003-4259-1972)，硕士研究生，E-mail：wangzhenmeng@mail.las.ac.cn。
收稿日期：2016-03-29 发表日期：2016-05-27 本文责任编辑：易飞



条数据基本上都是图书馆学与情报学（G25）和计算机科学（TP3）领域的文献，并通过文章第一作者所在的机构信息对此进行了验证。因此在英文数据的收集过程中，为了确保对比的有效性，本文对研究方向做了限定。本文中英文数据来源信息具体如表 1 所示：

表 1 中英文数据来源信息

数据	检索日期	来源数据库	检索结果（条）	检索式
中文数据	2016 年 1 月 2 日	中国学术期刊全文数据库	695	时间 (2006,2016) and (主题（精确）= 关联数据 or linked data or 关联开放数据 or linked open data or 链接数据 or 链接开放数据)
英文数据	2016 年 1 月 2 日	Web of Science Core Collection	1 314	time(2006,2016) and (subject=linked data or linked open data) and (language=English) and (研究方向 =computer science or information science & library science)

本文章主要应用统计分析、关键词共现分析和对比分析 3 种研究方法。统计分析主要是试图从数据量的角度来反映研究的发展现状以及未来的发展趋势；关键词共现分析的主要原理是通过关键词对共同出现的次数来反映二者之间的亲疏关系，进而反映相关研究主题之间的关系结构 [2]；对比分析可以很好地反映出比较对象之间的不同，在理论研究方面，比较结果可以很好地指导劣势方的发展。

3 数据分析

在正式的数据分析之前，本文对所获取的数据中的关键词项进行了基本清理，清理方式主要有中英文合并、单复数合并以及低价值词汇剔除等。为了对数据本身有更深入的理解，以下从发文量、词频、主题聚类及核心主题判定等方面对数据内容进行深度挖掘。

3.1 发文量分析

发文量的变化能够反映一个领域的发展状况以及未来的研究趋势，对国内外关联数据研究数据按照时间进行统计，其结果见图 1。关联数据的概念最早在 2006 年被提出 [3]，2008 年之后便进入研究的快速发展期，2014 年达到了发文量的峰值，之后的研究热度呈下降趋势。国际上对于关联数据的研究表现为理论与实践共同发展的态势，随着关联数据理论研究主题的丰富和内容的深入，国外关联数据事业也得到了长足的发展。相比于国际来说，国内相关研究具有一定的滞后性 —— 国内最早在 2008 年 12 月引入了关联数据这一概念，而最早见刊的

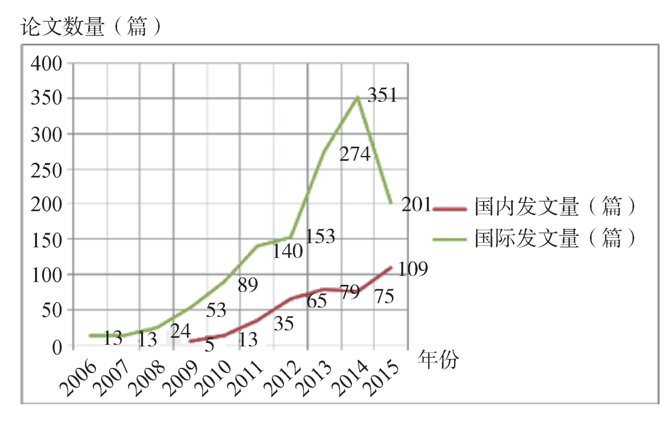


图 1 国内外关联数据研究论文数量走势

知识管理论坛，2016(3):163-173
DOI：10.13266/j.issn.2095-5472.2016.021

是姚小乐、刘炜等人翻译自 E. Summers 等人的《LCSH,SKOS 和关联数据》一文，该文从信息组织的角度对关联数据做了基本的介绍 [4]；国内关联数据研究的增长也较为迅速，并呈现进一步加速增长的趋势，预计 2017 年会出现发文量的峰值，但由于受到国际研究热度下降的影响，国内发文量的峰值也可能会提前到来。另外，笔者还根据分类号对文章的学科信息进行了统计，中文数据中计算机科学文献占比约为 51%（202 篇），图书馆学情报学文献占比约为 49%（196 篇），同属于这两个学科的文献共 17 篇；英文数据中计算机科学文献占比约 92%（1 275 篇），图书馆学情报学文献占比为 8% 左右（107 篇），同属于两个学科的文献共 68 篇。

3.2 词频统计分析

关键词是文章研究内容的直接体现，词频的高低可以体现出关键词之间重要性的差异；对国内外相关数据进行关键词统计及共词分析能够反映关联数据研究的热点主题以及热点主题之间的关系结构。本文章使用 SATI 3.2 进行关键词抽取和词频统计，中文数据中共含 802 个关键词，英文数据共含 2 420 个关键词。由于篇幅所限，这里只列出词频大于或等于 4 次的中文关键词和词频大于或等于 8 次的英文关键词（见表 2），其中省略了核心关键词“关联数据（linked data）”和“关联开放数据（linked open data）”。另外，为了更好地展现高频关键词之间的关系，本文分别采用关键词共现矩阵和关键词相似度矩阵两种方式对中英文数据进行展示；关键词共现矩阵可以让人直观地了解到两两关键词之间的共现频次以及二者之间的亲疏关系，而关键词相似度矩阵是在关键词共现矩阵的基础上为了减小误差进行的归一化处理，相似度矩阵中的数值都在 0-1 之间，数值越接近 1，表示两个关键词之间的主题相似度越大，反之越小。本文分别利用 Bicomb 2 和 SATI 3.2 构建了两类矩阵，但由于篇幅所限，文章只展示英文关键词词频大于 20 的关键词共现矩阵（12*12 矩阵，见表 3）以及中文关键词词频大于 10 的关键词相似度矩阵（11*11 矩阵，见表 4）。

3.3 主题聚类及主题识别

主题聚类是文献计量学常用的研究方法，已被广泛应用于某一领域的主题分析等方面，它利用词与词之间的关联程度进行聚类，形成多个主题。本文对中文关键词词频大于或等于 5 次的数据进行聚类分析，去除核心关键词“关联数据”和“关联开放数据”后共计 29 个（占关键词总数约 4% 的比例）。之后在前文分析的基础上，选择 SPSS 中的 ward 聚类算法进行系统聚类，并得到聚类树状图（见图 2）。之所以考虑 ward 算法，是因为该算法强调簇内的差异小、簇间的差异大，突出强调主题簇内部的同质性，适用于这种主题相关的数据样本进行进一步集群聚类。以选取的中文数据样本为参照，选取前 4% 的英文关键词（约 100 个关键词，词频亦为大于等于 5 次），由于 SPSS 系统聚类所能处理数据量有限，对于英文数据的主题聚类文章采用 Citespace 进行关键词聚类，在未做人为调整的前提下，数据的聚类效果见图 3。主题聚类的直接目的在于主题识别，主题识别的精准与否与前期的主题聚类效果有着重要的关系。本阶段的主题识别主要是通过关键词反向回溯文献再进行人为识别，所识别的结果直接来源于分析数据。由于所有关键词之间并不存在严格的层级关系，在忽略树状图本身所具有的层级特性的前提下，共识别出 7 个热点研究主题，同时在前文关键词共现矩阵的基础上，通过计算各主题的战略坐标来显示研究主题的热度和成熟度，热点研究主题依次为关联数据、本体与信息组织（E）、关联数据发布（C）、关联数据与图书馆服务（D）、关联数据知识发现（F）、关联数据资源整合（G）、关联数据可视化（B）、关联数据语义互联（A）。

英文数据的主题识别是基于 Citespace 所形成的 20 个类团图进行的，去除 8 个没有形成明显聚类的孤立点后，共形成了 12 个主题类团（见图 3），类团的位置可以反映研究主题所处的位



置，同时由于自动生成的类团标签并未能清晰反映研究主题，本文通过对类团内关键词进行人工回溯的方法形成了新的类团标签，表 5 在给出 12 个主题具体标签的同时又给出了对应的关键词，需要说明的是所形成的主题标签具有一定的主观性，相互之间也并非是完全独立和平等，但所形成的标签基本可以客观反映类团的核心研究主题。

表 2 关联数据研究文献关键词词频数据

中文关键词	词频	英文关键词	词频	中文关键词	词频	英文关键词	词频
语义网	62	semantic web	248	开放数据	5	natural language processing	10
RDF	35	ontology	115	信息服务	5	information extraction	10
图书馆	33	RDF	93	语义互联	5	data quality	10
本体	31	SPARQL	49	信息资源	5	integration	9
数字图书馆	15	DBpedia	27	知识链接	4	reasoning	9
书目数据	14	web of data	26	高校图书馆	4	data mining	9
知识组织	13	metadata	25	数据融合	4	learning	9
元数据	12	open data	24	数据集	4	cloud computing	9
数据网络	12	data integration	23	语义关联	4	data management	9
URI	8	provenance	21	知识组织系统	4	knowledge management	9
D2R	8	OWL	17	语义网技术	4	management	9
数据可视化	8	semantic annotation	17	机构知识库	4	mapping	8
知识服务	8	web	17	农业	4	vocabularies	8
SPARQL	7	information retrieval	15	数据共享	4	evaluation	8
SKOS	6	verification	14	SPARQL 查询	4	question answering	8
MARC	6	SKOS	14	信息技术	4	open government data	8
数据发布	6	interoperability	14	DBpedia	4	internet of things	8
信息聚合	6	big data	13	RDA	4	crowdsourcing	8
知识发现	6	visualization	12	元数据标准	4		
资源整合	6	wikipedia	11	信息组织	4	knowledge representation	8
OCLC	6	semantic search	11	科学数据	4	web services	8
Drupal	5	languages	10	FRBR	4	similarity	8
语义 Web	5	social media	10	数字档案资源	4		
数字资源	5	cultural heritage	10				
共词分析	5	interlinking	10				

知识管理论坛，2016(3):163–173
DOI：10.13266/j.issn.2095–5472.2016.021

表 3 英文高频关键词共现矩阵（部分）

关键词	linked data	semantic web	linked open data	ontology	RDF	SPARQL	DBpedia	web of data	metadata	open data	data integration	provenance
linked data	477	158	6	64	49	27	14	20	9	11	9	16
semantic web	158	248	31	37	25	20	8	11	9	4	3	7
linked open data	6	31	118	13	10	5	3	3	3	3	4	0
ontology	64	37	13	115	9	9	0	0	4	3	2	1
RDF	49	25	10	9	93	23	3	3	3	2	2	0
SPARQL	27	20	5	9	23	49	2	2	0	0	1	1
DBpedia	14	8	3	0	3	2	27	0	0	1	0	1
web of data	20	11	3	0	3	2	0	26	0	0	0	1
metadata	9	9	3	4	3	0	0	0	25	1	0	0
open data	11	4	3	3	2	0	1	0	1	24	0	0
data integration	9	3	4	2	2	1	0	0	0	0	23	0
provenance	16	7	0	1	0	1	1	1	0	0	0	21

表 4 中文高频关键词相似度矩阵（部分）

关键词	关联数据	语义网	RDF	图书馆	本体	数字图书馆	书目数据	知识组织	元数据	关联开放数据	数据网络
关联数据	1	0.139	0.058 4	0.081 9	0.044 5	0.014 7	0.041 6	0.0448	0.014 1	0.004 6	0.028 7
语义网	0.139	1	0.023 2	0.012 2	0.033 3	0.009 7	0	0.005	0.005 4	0.033 6	0.048 4
RDF	0.058 4	0.023 2	1	0.003 6	0.003 8	0	0.008 4	0	0.009 8	0.002 5	0.002 5
图书馆	0.081 9	0.012 2	0.003 6	1	0.003 9	0	0	0.009 3	0.002 5	0	0.022 7
本体	0.044 5	0.033 3	0.003 8	0.003 9	1	0.008 6	0	0.009 9	0.002 7	0.010 8	0
数字图书馆	0.014 7	0.009 7	0	0	0.008 6	1	0.004 8	0.020 5	0	0	0
书目数据	0.041 6	0	0.008 4	0	0	0.004 8	1	0.005 5	0	0	0
知识组织	0.044 8	0.005	0	0.009 3	0.009 9	0.020 5	0.005 5	1	0	0	0
元数据	0.014 1	0.005 4	0.009 8	0.002 5	0.002 7	0	0	0	1	0	0
关联开放数据	0.004 6	0.033 6	0.002 5	0	0.010 8	0	0	0	0	1	0.027 8
数据网络	0.028 7	0.048 4	0.002 5	0.022 7	0	0	0	0	0	0.027 8	1

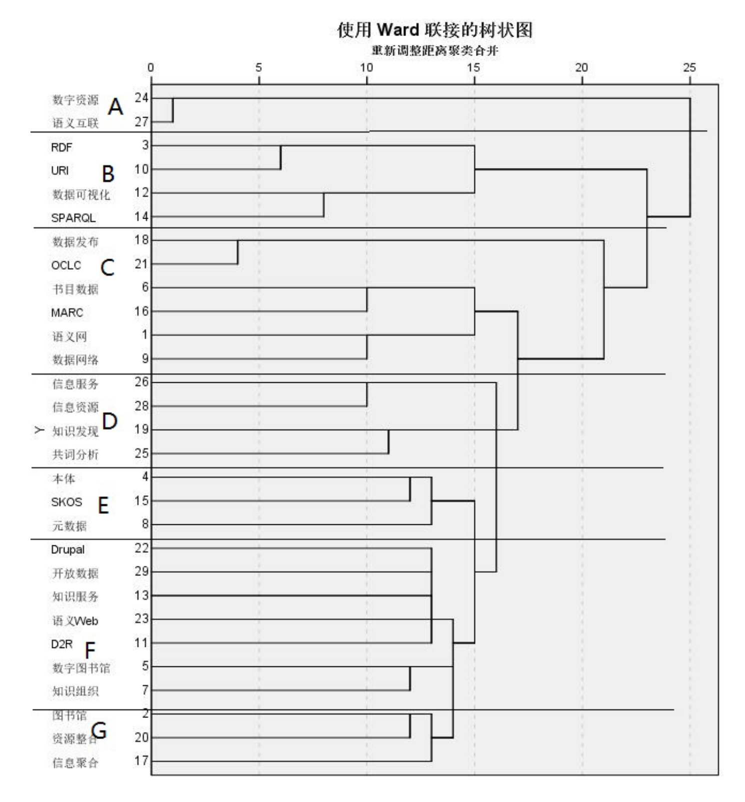


图 2 中文关键词 SPSS 聚类

表 5 英文数据主题聚类分析数据

序号	主题标签	对应关键词（部分）	序号	主题标签	对应关键词
1	本体、词表与元数据 [5]	ontology, ontology matching, disambiguation, lexicon, metadata, lemon	7	关联数据质量问题 [11]	description logics, streams, data quality, record linkage, extraction, biology
2	情感分析与语义挖掘 [6]	semantic, social media, opinion mining, sentiment analysis, classification	8	大型关联开放数据的可视化 [12]	big data, neural network, open data, visualization, models, ontology management, mapreduce, big and open linked data(BOLD)
3	关联数据管理与关联发现系统、工具 [7]	system, software, management, design, performance, algorithms, discovery, provenance	9	SPARQL 查询语言与关联数据在政治、教育和商业中的应用 [13]	SPARQL, Bayesian networks, reuse, open government, mooc, e-commerce, usability, information, impact
4	关联数据服务分析 [8]	ontology learning, services, service composition, ranking, data mining, framework, privacy	10	DBpedia 数据集质量评估 [14]	DBpedia, evaluation, testing, specification, semantic annotation, Wikipedia, domain ontology, patterns
5	语义检索 [9]	query language, search engine, semantic search, information retrieval, similarity, bootstrapping, sparql	11	关联数据发布与资源整合 [15]	data integration, information integration, natural languages, biomedical ontology, data publishing, science
6	知识库 [10]	semantic web, data transformation, knowledgebase, database, tool, information extraction, statistics, resource	12	关联数据推荐系统与隐私保护 [16]	social networks, recommender systems, machine learning, security, identity, protection

ChinaXiv合作期刊
知识管理论坛，2016(3):163–173
DOI：10.13266/j.issn.2095–5472.2016.021

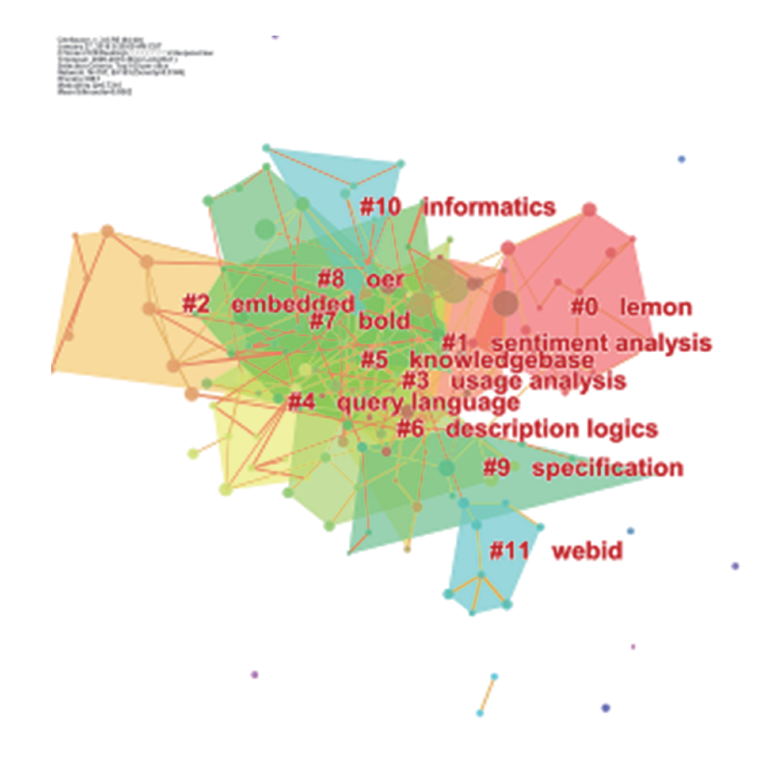


图 3 英文关键词 Citespace 聚类

4 中外关联数据热点研究对比分析

对比分析是认识事物并发现比较对象间不同点的有效方法，通过中外关联数据研究之间的对比不仅可以发现国内外关联数据研究存在的差异，同时也可以指导国内关联数据研究的未来发展方向。

4.1 研究群体的差异

中外关联数据研究群体存在较大的差异，这在一定程度上也导致了二者在研究主题方面存在的差异。研究群体的差异本质上是一种研究视角和研究方法的差异，不同的研究群体从各自学科的角度来审视同一研究主题，即使是具有一定交叉的某些学科，其所关注的研究点以及研究成果也天然带有某学科的学科特性，这主要是由研究人员的知识结构及其受过的学术训练决定的。从国内外关联数据的研究文献数据来看，承载关联数据这一研究课题任务的学科主要为计算机科学和图书馆学情报科学，因而国内外关联数据研究群体的不同主要体现在这两个学科在关联数据主题上比例的不同，比例的不同又可以从文献所属学科和研究人员单位两个角度来反映。从文献所属学科的角度来说，国外图书情报科学文献量占调研数据的 8%，而国内的比例则高达近 50%；从研究人员所属的研究单位来看，国内研究关联数据的学者中来自图书情报领域者比例高达 90% 以上，而国外则不到 10%。除此之外，尽管研究人员的数量、核心作者数量以及单个作者的发文量等数据都可以反映研究群体的差异，但由于文章是将国内数据与国外多个国家的数据进行对比，绝对数量的对比具有一定的不合理性，故文章未对此进行论述。总体来说，中外关联数据的研究群体在构成比例方面存在较大的差异，国内研究群体以来自图书馆学情报领域的研究人员为主，国外主要的研究力量则来自于计算机领域。



知识管理论坛
2016 年 第 3 期(总第 3 期)

4.2 研究主题的广度差异

国外关联数据研究主题较国内全面、完善，并逐渐形成了较为完整的关联数据研究体系，而国内相关研究依旧存在较多的空白点。对照 3.3 部分的热点研究主题发现，国外关联数据研究基本覆盖了关联数据生命周期（见图 4）的各个阶段，而国内的关联数据研究主题主要集中于关联数据前期研究的部分阶段，其余部分阶段的研究则基本空白或刚刚起步。对比中外关联数据热点研究主题，可以发现国内研究的空白点和新兴研究点（见表 6）——国外目前的研究热点中国内尚未出现的研究主题属于国内空白点，可为国内关联数据研究中长期的发展提供方向；国外目前的研究热点中国内已经出现但尚未发展成为热点的研究主题属于新兴研究点，可为国内关联数据当下的研究方向提供参考。

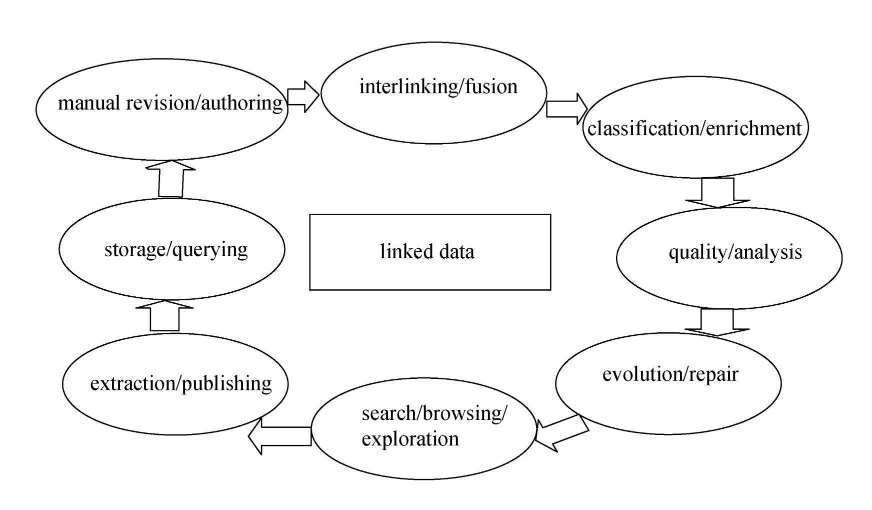


图 4 关联数据生命周期

表 6 国内关联数据研究中的新兴研究点与研究空白点

新兴研究点	研究空白点
关联数据术语服务 [17]	关联数据本体建模
关联数据挖掘 [18]	关联数据本体匹配、融合
关联发现 [19]	关联数据情感分析
关联数据权益保护 [20]	关联数据观点挖掘
关联数据搜索引擎 [21]	关联数据管理系统
SPARQL 查询语言 [22]	关联数据算法研究
关联数据质量问题 [23]	关联数据本体学习
关联数据推荐系统 [24]	关联数据语义标注
关联数据可视化 [25]	关联数据信息抽取
关联数据与电子政务 [26]	关联数据知识库
关联数据与教育资源共享 [27]	关联数据与电子商务
关联数据集质量评估 [28]	关联数据机器学习
关联数据机构评价 [29]	关联数据的学科应用
关联数据资源整合 [30]	关联数据自然语言处理
	关联数据信息安全

4.3 研究主题的深度差异

由 3.3 部分 SPSS 分析得到国内关联数据研究主要集中于 7 个研究主题，国外关于这些问题的研究已经较为成熟或起步较早，但即使对于相同的研究问题，国内外研究也存在较大的不同，即针对同一研究主题的研究深度亦有不同。主要体现在两个方面：一是同一主题下的子研究主题的多样性有所差异，二是同一研究主题的应用性研究数量不同。以关联数据生命周期的第一个阶段，同时也是国内目前研究相对较为成熟的“关联数据发布”研究主题为例，其国内外研究差异如表 7 所示：

4.4 研究主题的发展趋势及研究模式的差异

国内外关联数据研究不仅在当前具有较大差异，甚至在未来的研究发展方向和趋势方面也会存在一定的差异。国内关联数据研究的发展方向主要为关联数据可视化、关联数据挖掘

知识管理论坛，2016(3):163–173
DOI：10.13266/j.issn.2095–5472.2016.021

表 7 国内外关于“关联数据发布”的研究对比

分析点	国内“关联数据发布”研究	国外“关联数据发布”研究
研究多样性	关联数据发布工具研究[31] 图书馆关联数据发布方法研究 政府关联数据发布[26] 图书馆书目数据发布 图书馆词表资源发布	关联数据发布系统、工具开发 艺术、环境、农业、教育、政府、医药等多领域的关联数据发布 书目、词表、本体、元数据、科学数据的关联数据发布 关联数据发布标准、方法研究
研究类型	国内外研究综述、国外案例分析、理论研究	实证性研究、学科应用研究、理论研究、系统开发研究、案例研究
特点	主要为现状综述和理论研究，鲜有应用性研究，论述对象主要为图书馆馆藏资源	主要为学科领域应用性研究，基本都是实证性研究，理论研究与应用研究协调发展

等新兴研究点和研究空白点，且短期内研究的主要方式仍是介绍国外相关主题的研究现状和内容，但随着国内研究的深入，各主题领域的应用性研究数量可能会逐渐增多。相较而言，国外关联数据质量研究体系更为完善，理论与应用研究协调发展，研究已经逐步进入了缓慢发展时期，研究数量可能会继 2015 年后进一步减少；关联数据整合应用和学科化应用和关联数据联合查询以及关联数据质量问题是目前国外关联数据研究遇到的瓶颈问题，这也是国外研究的主要发展方向。

整体上来看，国内外关联数据研究的模式存在较大的差异，国外研究进程基本属于循序渐进的模式，在理论研究与应用研究协调发展的基础上，稳步地推进并理顺关联数据生命周期中各阶段的循环发展，是一种需求驱动的发展形式。国内则可能是由于起步时间较晚以及研究人员的学科背景差异，在对国外研究的重点介绍关联数据研究中占有很大的比重，研究形式多以综述和案例分析为主，同时也并非是依据研究需求的发展依次引入关联数据生命周期中各阶段的研究主题，例如在关联数据前期各阶段研究尚未成熟的情况下，关联数据整合已成为国内关联数据研究的热点主题。

5 结论

通过对国内外关联数据研究领域的发文量、词频、主题、研究群体等的分析，不仅可以发现国内外相关研究存在一些差异，也能指导国内关联数据研究的发展：首先，可以优化国内研究群体结构，引入新的研究力量——学科背景不同在一定程度上决定了研究关注的重点，研究形式等也会有所差异，引入计算机科学领域的研究力量是深化国内关联数据研究的途径之一，有助于关联数据研究的合理、健康发展。其次，在深化理论研究的同时，应积极开展应用型研究。即不仅要关注关联数据概念、原则、发布方法、整合理论等方面的理论研究，同时更要关注发布工具、资源整合工具以及质量评估工具的开发应用。再次，应关注关联数据的学科化应用，打破图书馆概念实体的限制。在关注图书馆关联数据应用发展的同时，更应积极探索关联数据在农业、环境、教育、医药等领域的专业化应用。最后，要注重关联数据研究的本土化。国内当前的关联数据研究所基于的数据资源基本都是外文资源，由于中文数据在转化为关联数据的过程中会碰到一些中文数据所特有的问题，如 URI 配置等，关联数据研究本土化是国内关联数据事业发展的首要前提之一。

参考文献：

[1] 刘炜 . 关联数据：概念、技术及应用展望 [J]. 大学图书馆学报 , 2011, 29(2): 5-12.
[2] 冯璐 , 冷伏海 . 共词分析方法理论进展 [J]. 中国图书馆学报 , 2006, 32(2): 88-92.
[3] BERNERS-LEE T. Linked data [EB/OL]. [2016-03-06].

https://www.w3.org/DesignIssues/LinkedData.html.
[4] SUMMERS E, ISAAC A, REDDING C, et al. LCSH, SKOS 和关联数据 [J]. 姚小乐，刘炜译 . 现代图书情报技术，2009(3): 8-14.
[5] SCHARFFE F, EUZENAT J. Linked data meets ontology matching: enhancing data linking through ontology alignments[C]//International conference on knowledge engineering and ontology development. Paris: Springer, 2011: 279-284.
[6] OSIMO D, MUREDDU F. Research challenge on opinion mining and sentiment analysis[J/OL]. [2016-03-20]. http://www.w3.org/2012/06/pmob/opinionmining.pdf.
[7] ZHAO J, MILES A, KLYNE G, et al. Linked data and provenance in biological data webs [J]. Briefings in bioinformatics, 2009, 10(2): 139-152.
[8] SPEISER S, HARTH A. Integrating linked data and services with linked data services [M]. Berlin: Springer, 2011: 170-184.
[9] HOGAN A, HARTH A, UMBRICH J, et al. Searching and browsing linked data with swse: the semantic web search engine [J]. Web semantics: science, services and agents on the world wide web, 2011, 9(4): 365-401.
[10] VRANDEČIĆ D, KRÖTZSCH M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
[11] MENDES P N, MÜHLEISEN H, BIZER C. Sieve: linked data quality assessment and fusion[C]. Proceedings of the 2012 joint EDBT/ICDT workshops. Berlin: ACM, 2012: 116-123.
[12] BRUNETTI J M, AUER S, GARCÍA R, et al. Formal linked data visualization model[C]//Proceedings of international conference on information integration and web-based applications & services. Sydney: ACM, 2013: 309.
[13] PIEDRA N, CHICAIZA J A, LÓPEZ J, et al. An architecture based on linked data technologies for the integration and reuse of OER in MOOCs context [J]. Open praxis, 2014, 6(2): 171-187.
[14] ZAVERI A, KONTOKOSTAS D, SHERIF M A, et al. User-driven quality evaluation of dbpedia[C]//Proceedings of the 9th international conference on semantic systems. Graz: ACM, 2013: 97-104.
[15] CORRENDO G, SALVADORES M, MILLARD I, et al. SPARQL query rewriting for implementing data integration over linked data[C]//Proceedings of the 2010 EDBT/ICDT workshops. Berlin: ACM, 2010: 4.
[16] DI NOIA T, MIRIZZI R, OSTUNI V C, et al. Linked open data to support content-based recommender systems[C]//Proceedings of the 8th international conference on semantic systems. New York: ACM, 2012: 1-8.
[17] 范炜 . 受控词表的术语服务研究 [J]. 图书情报工作，2012, 56(14): 34-39, 97.
[18] 刘炜，夏翠娟，张春景 . 大数据与关联数据：正在到来的数据技术革命 [J]. 现代图书情报技术，2013 (4): 2-9.
[19] 沈志宏，黎建辉，张晓林 . 关联数据互联技术研究综述：应用，方法与框架 [J]. 图书情报工作，2013, 57(14): 125-133.
[20] 张春景，刘炜，夏翠娟 . 关联数据的开放应用协议 [J]. 中国图书馆学报，2012, (1): 43-48.
[21] 黄永文 . 关联数据在图书馆中的应用研究综述 [J]. 现代图书情报技术，2010, 26(5): 1-7.
[22] 汤怡洁，杨锐，刘毅 . 近年来 SPARQL 查询技术的研究热点及进展 [J]. 知识管理论坛，2014(1): 1-6.
[23] 郭少友，尚雨，魏朋争，等 . 绿色关联数据：概念、要求与工程 [J]. 图书情报工作，2012, 56(23): 113-118.
[24] 田野，祝忠明，刘树栋 . 基于关联数据的推荐系统综述 [J]. 现代图书情报技术，2013, 29(10): 1-7.
[25] 陈涛，夏翠娟，刘炜，等 . 关联数据的可视化技术研究与实现 [J]. 图书情报工作，2015, 59(17): 113-119.
[26] 吴玥，李占羽 . 基于关联数据开放政府数据 [J]. 电脑知识与技术，2010 (11): 8688-8691.
[27] 吴鹏飞，马凤娟 . 国外关联数据教育应用项目研究与启示 [J]. 电化教育研究，2013, 34(4): 114-120.
[28] 魏来，付瑶 . 基于 Green Linked Data 的关联数据质量标准 [J]. 情报资料工作，2013, 34(3): 69-72.
[29] 蒋世银，李春旺 . 基于关联数据科研机构综合评价模型研究 [J]. 情报理论与实践，2015, 38(11): 71-75.
[30] 欧石燕，胡珊，张帅 . 本体与关联数据驱动的图书馆信息资源语义整合方法及其测评 [J]. 图书情报工作，2014, 58(2): 5-13.
[31] 白海燕，梁冰 . 利用 D2R 实现关系数据库与关联数据的语义模式映射 [J]. 现代图书情报技术，2011(8): 1-7.

知识管理论坛，2016(3):163–173
DOI：10.13266/j.issn.2095–5472.2016.021

A Contrastive Analysis of Linked Data Research Hotspots at Home and Abroad

Wang Zhenmeng[1,2]
[1]University of Chinese Academy of Sciences, Beijing 100049
[2]Chengdu Documentation and Information Center, CAS, Chengdu 610041

Abstract: The same to Chinese abstract. **[Purpose/significance]** The domestic research on linked data is still in its initial stage. With theoretical research and application research being immature, this study could provide some inspiration and references fordomestic research on linked data. **[Method/process]** By reviewing lots of literature and the systemic analysis, this study has analyzed the differences between domestic and international researches in terms ofthe quantity of published papers, research groups, word frequency and subject identification. **[Result/conclusion]** Based on the analysis above, this article proposes four tips to promote the development of domestic research on linked data, including indigenous research, paying more attention to application research, optimized the structure of research groups and focusing on multidisciplinary applications.

Keywords: linked data　researchhotspots　contrastive analysis　data analysis